ENVIRONMENTAL RESEARCH
LETTERS

LETTER • OPEN ACCESS

Drinking water quality and social vulnerability linkages at the system level in the United States

To cite this article: Bridget R Scanlon *et al* 2023 *Environ. Res. Lett.* **18** 094039

IOP Publishing Environ. Res. Lett. 18 (2023) 094039 https://doi.org/10.1088/1748-9326/ace2d9

ENVIRONMENTAL RESEARCH
LETTERS

OPEN ACCESS

RECEIVED
2 February 2023

REVISED
1 June 2023

ACCEPTED FOR PUBLICATION
29 June 2023

PUBLISHED
5 September 2023

LETTER

Drinking water quality and social vulnerability linkages at the system level in the United States

Bridget R Scanlon[1,*], **Robert C Reedy**[1], **Sarah Fakhreddine**[1,2], **Qian Yang**[1] **and Gregory Pierce**[3]

[1] Bureau of Economic Geology, Jackson School of Geosciences, University of Texas at Austin, Austin, TX, United States of America
[2] Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA, United States of America
[3] Luskin Center for Innovation, University California Los Angeles, Los Angeles, CA, United States of America
[*] Author to whom any correspondence should be addressed.

E-mail: bridget.scanlon@beg.utexas.edu

Keywords: drinking water quality, Safe Drinking Water Act, community water system boundaries, social vulnerability, Bipartisan Infrastructure Law, disadvantaged communities

Supplementary material for this article is available online

Abstract

Growing recognition of water quality concerns, particularly in socially vulnerable communities in the United States, has prompted recent policies and investments to improve drinking water system performance. Current environmental justice tools limit measurement of drinking water quality issues to proximity to point-source contamination, such as superfund sites and social vulnerability to county level or zip code level data. We examined relationships between health-based (HB) drinking water quality violations and social vulnerability using a new database of community water system (CWS) service areas and a modified Social Vulnerability Index, which we specifically designed for drinking water quality. CWSs with HB violations disproportionately impact socially vulnerable communities, with ∼70% of such systems characterized by high social vulnerability. Increased risks of drinking water quality violations in high socially vulnerable communities are attributed in part to violations related to pervasive, naturally occurring contaminants (e.g. arsenic) requiring treatment and difficulties for small systems to implement and maintain treatment systems. Notably, recurrence of any HB violation is also related to social vulnerability ($R = 0.73$). The relative importance of different social parameters, including socioeconomic status, race and language, and demographics and housing characteristics, varies with the type of violation. Further understanding linkages between drinking water quality violations and social vulnerability is essential for optimizing the deployment of, and motivating the next tranche of newly available drinking water infrastructure funding that is heavily prioritized toward disadvantaged communities.

1. Introduction

While the vast majority of Americans have access to safe drinking water (table 1), providing high quality water is increasingly challenging, particularly in socially vulnerable communities [1–3]. Recently, the US federal government declared a state of emergency in Jackson, Mississippi, after severe storms exacerbated an ongoing drinking water crisis [4, 5], resulting in a federal civil rights investigation to determine whether state agencies responsible for funding water infrastructure discriminated against Jackson whose residents are 82% black and 25% in poverty [6]. Similarly, the 2014 crisis in Flint, Michigan, not only threatened public health, but also increased mistrust of public water supplies [7] and tap water avoidance in Black and Hispanic households across the US [8]. Nationally, a higher proportion of minority households perceive tap water to be unsafe and subsequently purchase bottled water, which costs ∼2000 times more than tap water [9].

In addition to local infrastructure failures such as Jackson and Flint, there is growing concern over the general state of US drinking water infrastructure, which is known to be aging and underfunded [10, 11]. Recent state and federal initiatives have focused on improving water infrastructure and addressing disparities in access to safe, affordable, and

Table 1. List of selected acronyms. A more detailed list of acronyms is provided in supporting information, section 1.

ACS	American Community Survey
CDC	Centers for Disease Control
CEJST	Climate and Economic Justice Tool
CWS	Community water system
DAC	Disadvantaged community
DBP	Disinfectants and disinfection byproducts
DW	Drinking water
GW	Groundwater
HB	Health-based
RTC	Revised total coliform
SDWA	Safe Drinking Water Act
SDWIS	Safe Drinking Water Information System
SVI	Social Vulnerability Index
SW	Surface water

reliable drinking water. For example, several states have declared water a basic human right, including California, Massachusetts, and Pennsylvania [12, 13]. At the federal level, 2021 Bipartisan Infrastructure Law provides $50 billion over five years to address infrastructure challenges, including community water systems (CWSs) that do not comply with the EPA Safe Drinking Water Act (SDWA) [14, 15].

Recent national studies have evaluated drinking water quality violations in the US [16, 17]. Analysis of the dominant causes of SDWA health-based (HB) violations in CWSs shows that disinfectants and disinfection byproduct (DBP) violations related to water treatment were dominant [18], followed by naturally occurring contaminants (e.g. arsenic and radionuclides) and non-point source anthropogenic contaminants, especially nitrates [16]. Data analytics were used to determine dominant drivers of violations, including environmental (landcover, climate, geology), operational (water source, system size), and sociodemographic (social vulnerability, rurality) drivers. The Social Vulnerability Index (SVI) developed by the Center for Disease Control (CDC) [19, 20] was used in this study [16]. The CDC SVI was developed to address social vulnerability to natural (e.g. floods) or human caused disasters or disease outbreaks but not drinking water quality issues.

Despite growing recognition of U.S. drinking water quality inequities, emerging environmental justice tools remain vague on drinking water quality. For instance, the White House Climate and Economic Justice Tool (CEJST), designed to apply 40% of certain federal investments to disadvantaged communities (DACs) [21], has a water category which is limited to proximity to point-source contamination, including superfund sites, wastewater discharges, and underground storage tanks. The White House CEJST builds on earlier efforts by the EPA, who first released the EJ Screening and Mapping tool in 2015 [22]. EJScreen's environmental indicators emphasize air quality, and indicators linked to water quality focus on proximity to point sources of contaminants, similar to the White House CEJST. These tools do not capture DBPs and naturally occurring contaminants which dominate drinking water violations.

Additionally, defining disadvantaged or marginalized communities is critical for federal EJ tools and for states because states are tasked with administering 49% of the Bipartisan Infrastructure Law funds toward DACs. However, state-specific definitions of DACs mostly focus on socioeconomic factors, particularly median household income (MHI) level, which is used by 49 out of 51 states [23] rather than environmental justice metrics (e.g. race).

Previous studies relating water quality violations to social parameters were often conducted at the county level [1, 3, 24, 25]; however, some studies show large differences in water quality and other equity outcomes by CWSs within the same counties [26]. Many social parameters, such as socioeconomics, race and language, and housing type and tenure, are not collected at spatial scales specifically relevant for CWSs. Additionally, data on CWS boundaries or service areas are also limited. Only recently in 2022, CWS service areas were compiled for 15 states that make the data publicly available and data analytics were used to estimate boundaries for many other systems [27]. This new dataset greatly enhances our ability to assess linkages between water quality and social vulnerability at more appropriate scales.

Accordingly, this study aims to address the following questions:

(1) Do drinking water quality violations disproportionately impact socially vulnerable populations at the system level?
(2) Which social vulnerability parameters have the greatest impact on the probability of CWSs having different major types of violations?
(3) Is recurrence of violations linked to a customized measure of local social vulnerability?

Unique aspects of this study include (1) use of a new dataset delineating CWS service areas rather than relying on county level data [16, 28], (2) considering all drinking water quality HB violation rules rather than focusing solely on an individual or subset of rule violations [18, 29–31] or point-source contaminants [22, 32]; (3) incorporating all CWS size categories as previous studies have excluded small CWSs serving $\leqslant$500 people[24] or $\leqslant$5000 [33], (4) evaluating CWSs throughout the CONUS rather than focusing on particular regions [34, 35], and (5) assessing linkages to various social vulnerability parameters relevant to drinking water rather than limiting to economic parameters [23] (figure 1). Further, we introduce a novel modified social vulnerability index (mSVI) specific to drinking water quality violations by using key parameters identified from previous studies and data analytics. Temporal analyses included evaluation of the recurrence of SDWA violations over a three-year period (2018–2020) and relationship to our mSVI. The comprehensive in-depth evaluation of drinking water violations and novel drinking water mSVI developed in this study can help inform an emerging US national water equity strategy, leveraging increasing funding opportunities available over the next five years through the federal Bipartisan Infrastructure Law.

2. Methods

2.1. Spatial location of Community Water Systems

The flow chart describes the main data inputs, data analyses, and outputs of the study (figure 1). We used a newly developed dataset of tiers of CWS service area boundaries developed by EPIC/SimpleLab [36] (supporting information, SI, section 2.8; figure 2). We used CWSs with Tier 1 data which are based on states which provide explicit boundaries (15 states) and Tier 2 data which are produced by a matching algorithm to link CWSs to boundaries of a town or city. We excluded Tier 3 data which are estimated using a statistical model and considered less accurate. The Tier 1 and 2 data totaled ∼28 128 CWSs, representing 58% of all CWSs covering 85% of CWS population served (263 million out of 308 million people) and distributed across ∼3000 out of 3143 counties in the contiguous US (CONUS). We do not have CWS boundaries for all tribal systems.

2.2. Safe Drinking Water Act violation data

Our analysis focused on HB violations of the EPA SDWA which fall into three categories (table S2)

(1) Maximum contaminant levels (MCLs),
(2) Maximum residual disinfectant levels, and
(3) Treatment technique [37].

In cases where contaminant quantification is technically or economically infeasible, treatment rules, such as the Groundwater Rule, Surface Water Treatment Rules, and Lead & Copper Rule have been established. Additional data on SDWA regulations along with the health effects are provided in SI, section 2.1. The groupings we used resulted in ten specific rule violations, as follows: (1) arsenic, (2) nitrates, (3) inorganics, (4) organics [volatile and semivolatile organic compounds], (5) radionuclides, (6) Revised Total Coliform Rule, (7) DBP Rules, (8) Lead & Copper Rule, (9) Surface Water Treatment Rules, and (10) Ground Water Rule. For some analyses, we aggregated rules: 'any HB violation' refers to all HB violation rules, 'any inorganic' includes arsenic, nitrate, and inorganics rule violations.

Violation data from 2018 to 2020 (12 quarters) were selected because EPA regulations were fairly stable during this time. Analysis focused on systems active on 15 April 2021 (time of Safe Drinking Water Information System [SDWIS] data download). The number of systems with HB violations was considered along with populations served by these CWSs. Populations served by CWSs are defined as very small (⩽500 people), small (501–3300), medium (3301–10 000), large (10 001–100 000), and very large (⩾100 000) (table S3). The majority of CWSs (81%) are very small to small systems. The majority of CWSs are sourced by groundwater (72%), and are generally small, but most of the population (73%) is served by systems sourced by surface water.

2.3. Modified social vulnerability index for drinking water quality

Alongside the drinking water quality violation data, we modified the CDC SVI that was developed for natural hazards to focus on social vulnerability parameters that have either been identified in the literature to relate to disparities in drinking water quality or ranked highly in our preliminary random forest classification modeling (section 2.5). Differences between mSVI (mSVI, 15 US Census variables grouped into three themes) and the original CDC SVI are described in SI, section 2.4 (figure 3). These differences include expansion of the Socioeconomic Status theme to include per-capita income in mSVI because most states define DACs using income data [23]. Data on household composition were excluded from mSVI because attributes related to household composition were considered more relevant to hazard response. Minority status was expanded in mSVI to include Hispanic, Black, and Asian because many previous studies found linkages between drinking water quality noncompliance and ethnicity [1, 25, 28]. Most of the attributes related to housing in mSVI were retained from the CDC SVI but transportation 'no vehicle' was excluded because transportation is more important for hazard response. Population density was added to the mSVI as a proxy for rurality because previous studies found rurality to be important for drinking water quality violations [24].

Tract-level data were used in this analysis rather than higher resolution block-level data because the margins of error associated with blocks are considered too large. Tract level census variables were area weighted to CWS service areas. Following CDC methodology, we calculated mSVI using a percent rank across all census tracts resulting in normalized mSVI values ranging from 0 to 1. All tribal systems are currently classified as disadvantaged according to the EPA and therefore were assumed to have high levels of social vulnerability for the purposes of this study.

2.4. Data analytics to assess linkages between social vulnerability and health-based violation

We used balanced random forest (BRF) to assess linkages between social vulnerability parameters described in the previous section and occurrence of a

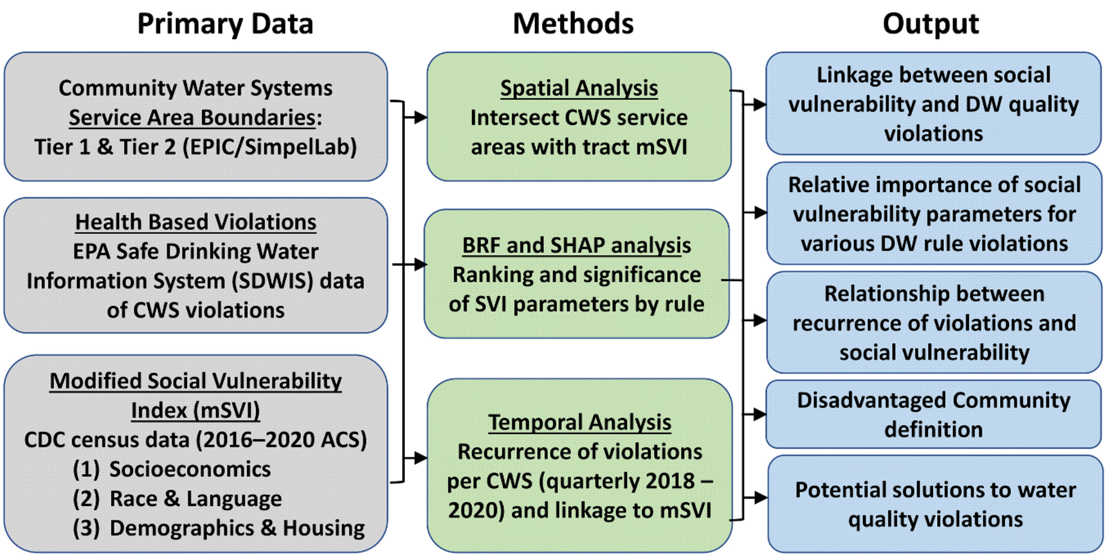

Figure 1. Flowchart showing data sources, spatiotemporal analysis and data analytics, and important study outputs. DW refers to Drinking Water.

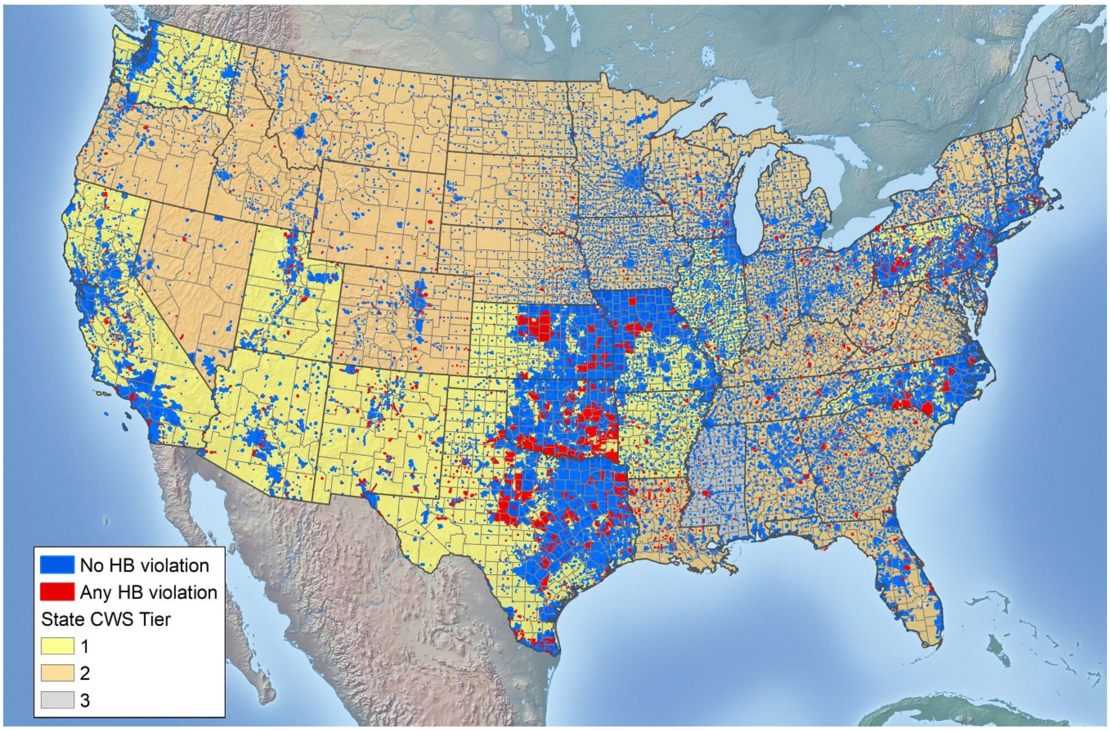

Figure 2. Boundaries for Tier 1 & 2 community water system service areas totaling 28 128 in the CONUS. Service area boundaries have been exaggerated as most are generally too small to be seen at this scale. State areas are highlighted by tier with Tier 1 representing 15 states that have published boundaries and Tier 2 representing states that have water system names that match census place names. Tier 3 system service areas are approximated and were excluded from our analysis.

HB SDWA violation, which is a binary response variable (±HB violation) (table S3). Because HB violations represent a highly imbalanced dataset with only ~10% of all CWSs in violation [16], BRF was selected because it is suitable for dealing with imbalanced datasets [38]. Details of the approach are provided in SI, section 2.5(b). The BRF analysis was based on occurrence of any single SDWA violation during the period 2018–2020, excluding multiple violations of the same rule from the model. Recurrence of violations over time was analyzed separately (section 2.5).

Output from BRF was used with SHapley Additive exPlanations (SHAPs) to assess the relative contribution of explanatory variables (social parameters) on the probability of a CWS having a violation. TreeSHAP, a variant of SHAP for tree-based models, was used to estimate SHAP values for tree models as it can handle dependent features [39, 40].

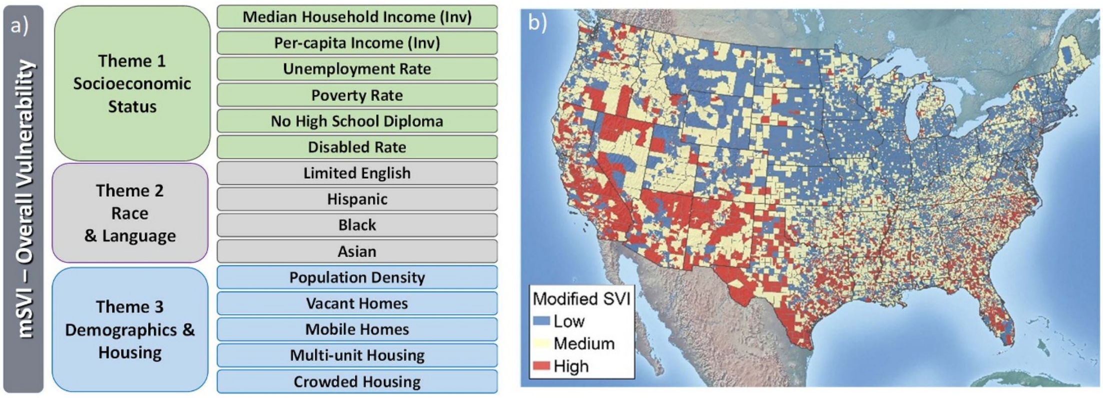

Figure 3. Modified Social Vulnerability Index (mSVI) based on three themes including 15 parameters and map of mSVI by US Census Tracts based on US Census Bureau (USCB) American Community Survey 5 yr data for 2016–2020. CWS data were created as the population-weighted sums of tract level mSVI values. Inv refers to inverted. Data are provided in table S4. Details are provided in SI, section 2.4 and mSVI is compared to the CDC SVI in figure S3.

Feature SHAP values represent the impact of the feature (social parameters) on the probability of a HB violation. Purely random data would have a probability of a violation of 0.5. SHAP values represent the absolute change in probability beyond 0.5 associated with each explanatory variable. Model output data, including the testing datasets, SHAP rankings, and selected partial dependence plots, can be found in tables S9–S19.

2.5. Analysis of temporal variability in Safe Drinking Water Act violations

We further examined temporal variability in HB violations, focusing on recurrence of violations based on the total number of quarters in violation from 2018 to 2020 (total of 12 quarters). Violations of some SDWA rules did not typically recur, including Surface Water Treatment Rule, Ground Water Rule, Revised Total Coliform Rule, and Lead & Copper Rule (table S21). For recurring violations, we calculated the relationship between recurrence and median mSVI of CWSs in violation.

3. Results and discussion

3.1. General distribution of drinking water quality violations

Approximately one in ten people were exposed to any HB violation (∼26 million out of 260 million people, table 2) over the three-year period from 2018 through 2020, corresponding to 11% of all Tier 1 and Tier 2 CWSs considered (3165/28128). This percentage of CWSs with any HB violation is similar to that for all CWS across the CONUS (10%; ∼5000/∼49000 CWSs).

The dominant cause of violations (by percent of CWSs with HB violations) is DBPs during water treatment, resulting in DBP Rule violations (38%) (figure 4). The next highest category is 'any inorganic + radionuclides' (which we define as combined arsenic, nitrate, and inorganics plus radionuclide rule violations) accounting for 22% of CWS violations, primarily related to naturally occurring (e.g. arsenic, uranium) and anthropogenic (e.g. nitrate) contaminants. Violations of the Groundwater Rule account for 21% of CWS violations, Revised Total Coliform Rule for 12%, Surface Water Treatment Rule for 11%, and Lead and Copper Rule for 7%. Organic rule violations account for the smallest fraction of violations (0.6% of CWSs) which is surprising considering that organic violations include an extensive list of 53 chemicals. Additionally, environmental justice issues of water quality often highlight community proximity to industrial facilities and superfund sites [28] rather than formation of DBPs and naturally occurring inorganics and radionuclides.

3.2. Modified social vulnerabiity index and linkage to drinking water quality violations

The mSVI map shows high mSVI in census tracts throughout the southern US and many areas in the western US. Low mSVI is shown in much of the Midwest and Northeast (figure 3(b)). Drinking water quality violations disproportionately impact socially vulnerable populations with the majority (71%) of the population impacted by any HB violation residing in the high mSVI tercile of CWSs included in this study (∼28 000 CWSs), whereas much lower populations fall into the middle (19%) and lowest (10%) terciles of mSVI (table 3). Therefore, ∼7 out of 10 people experiencing any HB violation are classified by high mSVI.

The relationship between population impacted and mSVI tercile varies by rule; 61% of the population impacted by arsenic rule violations plot within the high mSVI tercile, 68% for DBP rules, 49% for nitrate, and 42% for radionuclides (table 3). Violations of the remaining rules (Surface Water

Table 2. Summary statistics of Tier 1 and 2 CWSs and all CWSs with health-based violations from 2018 to 2020. Data are provided in table S3.

Rule/Rule group	Number of CWSs	Population served (millions)	GW source		SW source	
			No. of CWSs	% of CWSs	No. of CWSs	% of CWSs
Tier 1 and Tier 2 CWSs						
All	28115	260.32	20 232	72.0	7878	28.0
Any health-based	3280	25.92	1982	60.4	1298	39.6
Arsenic	272	0.48	258	94.9	9	3.3
Nitrates	237	0.19	228	96.2	16	6.8
Radionuclides	204	0.50	195	95.6	8	3.9
DBPRs	1246	9.79	331	26.6	915	73.4
SWTRs	350	9.78	8	2.3	342	97.7
GWR	671	1.30	626	93.3	31	4.6
RTCR	397	2.02	317	79.8	86	21.7
LCR	228	2.83	164	71.9	60	26.3
Organics	19	0.12	13	68.4	6	31.6
All community water systems						
All	48699	307.9	37 350	76.7	11 324	23.3
Any health-based	5041	29.0	3244	6.7	1797	3.7

GW: ground water; SW: surface water; DBPRs: disinfectants and disinfection byproducts Rule 1 and Rule 2; Arsenic Rule, Nitrates Rule, Radionuclides Rule, GWR: Ground Water Rule, Any inorganic: Arsenic Rule, Nitrates Rule, or Inorganics Rule, RTCR: Revised Total Coliform Rule, SWTRs: Surface Water Treatment Rule, Long-Term 1 Treatment Rule, or Long-Term 2 Treatment Rule, LCR: Lead and Copper Rule, Organics: includes Volatile Organic Compounds Rule or Synthetic Organic Compounds Rule.

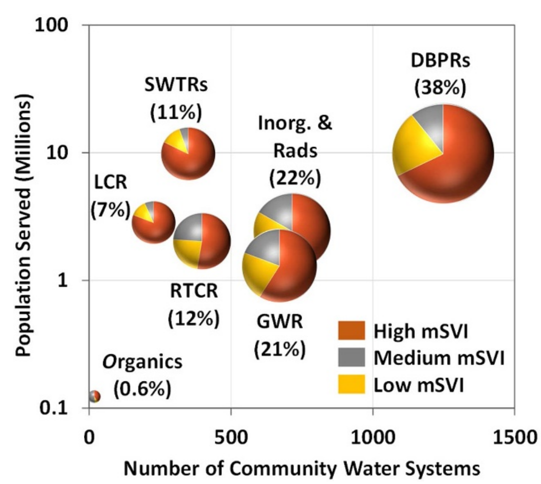

Figure 4. Comparison of population served by active community water systems (CWSs) with any HB violation in 2018–2020 relative to total number of active CWSs. Pie charts represent population served corresponding to mSVI terciles for low (0–0.33), medium (0.33–0.67), and high (0.67–1). Breakdown of CWSs with HB violations by rule was 38% Disinfectant and Disinfection Byproduct Rule (DBP Rules), 22% any inorganic including arsenic and nitrate and radionuclides (Inorg. & Rads), 21% groundwater rule (GWR), 12% revised total coliform rule (RTCR), 11% surface water treatment rule (SWTR), 7% lead & copper rule (LCR), and 0.6% organics. Percentages do not sum to 100% because some CWSs have multiple SDWA violations, which are counted separately in this analysis. Acronyms are listed in SI, section 1. Data are provided in table S3.

Treatment, Groundwater, Revised Total Coliform, Lead & Copper, and organics) correspond to 40%–82% of high mSVI populations.

Table 3. Distribution of populations according to modified Social Vulnerability Index (mSVI) terciles for different drinking water quality rule violations.

Violation type	% of population by mSVI Tercile		
	Low	Medium	High
Any HB	10.4	18.7	70.9
Arsenic	12.6	26.1	61.3
Nitrates	8.4	42.6	48.9
Rads	21.5	36.4	42.2
DBPR	10.6	21.7	67.7
SWTR	5.3	12.3	82.4
GWR	19.2	22.0	58.9
RTCR	23.9	23.4	52.7
LCR	6.6	12.5	80.9
Organics	52.8	7.6	39.6

HB: health based; Rads: radionuclides; DBPR: disinfectants and disinfection byproducts; SWTR: Surface Water Treatment Rule; GWR: Ground Water Rule; RTCR: Revised Total Coliform Rule; and LCR: Lead and Copper Rule.

Tribal CWSs were also disproportionately impacted by HB violations with ∼3 in 10 people impacted by a HB violation relative to 1 in 10 people impacted across the CONUS (figure 5). Tribal systems with violations are primarily located in Arizona, New Mexico, and California. In other words, a total of 18% of tribal CWSs had at least one HB violation, corresponding to 28% of the 1 million people served by tribal systems over 2018–2020. HB violations of tribal systems consisted primarily of the Ground Water Rule (44%), followed by Revised Total Coliform Rule (22%), and arsenic rule (16%) system

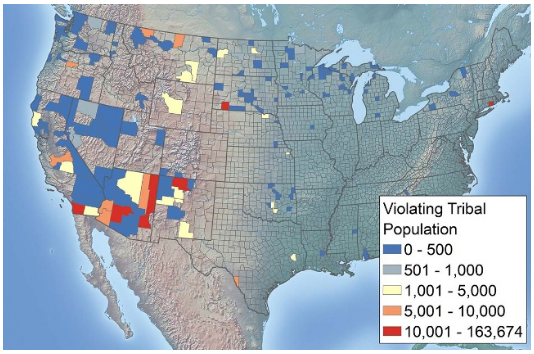

Figure 5. Map of population served by systems on tribal land served by CWSs with HB SDWA violations based on 2018–2020 data. Three out of ten people are susceptible to any HB violation in tribal CWSs. The analysis indicates that there was a total of 129 tribal CWSs (~18% of all CWSs serving native Americans) with any HB violation serving ~296 000 people (~28% of the ~1 million native Americans served by CWS) (table S20(e)). Only the Navajo nation has primacy.

violations (figure 5). These values may be underestimating HB violations due to deficiencies in SDWA enforcement in tribal systems [41].

3.3. Relative importance of different social vulnerability parameters in drinking water quality violations

Our analysis shows that many social parameters are linked to drinking water quality violations, with relative importance of different parameters varying by rule (figure 6). Socioeconomic parameters do not always rank most highly based on SHAP analyses, which contrasts with many state and federal tools relying almost exclusively on MHI to define DACs [23] or emphasizing socioeconomics as primary metrics for environmental justice disparities [22, 32].

Considering any HB violation, all three themes (socioeconomic, race + language, and demographics/housing) were represented in the top five social parameters in the SHAP results (figure 6(a)). Any HB violation was positively related to vacant homes (depopulation) increasing the probability of incurring a violation by 6.3%. The direction of the impact (positive or negative) is shown by the partial dependence plots (table S9(d)). Population density reduced the probability of any HB violation, also ranking high, and may reflect linkages between violations and rurality consistent with previous results[24]. Race and language were also important factors with Asian reducing the probability (4.6%) and Hispanic increasing the probability (2.8%) of any HB violation incurred by a CWS. These relationships may reflect Asian populations being more concentrated in urban settings and Hispanic population more broadly distributed throughout the southwest and southcentral US. By contrast, socioeconomic factors generally ranked lower, with per capita income ranking 5th, poverty ranking 8th, and MHI ranking 9th. Violations of arsenic, nitrate, any inorganic, and radionuclide rules exhibited a prominent linkage with race and language particularly with Hispanic population (positive impact, increasing probability of violation by 8.4%–18%) and limited English (positive, ~7%) within the top five parameters (figures 6(b), (c); tables S10–S13). Population density also ranked highly for these violations, with negative impacts on violation probabilities ranging from 6.7% to 14%. The predominance of these factors may reflect the geographic distribution of these violations, in mostly rural areas within the southwest and southcentral US [16]. Socioeconomic factors rank lower than race and language for these violations.

Socioeconomic factors are more prominently associated with DBP rule violations, with MHI, disabled, and no high school diploma positively related to probability of DBP violations (each ~6%) (figure 6(d)). Vacant housing was also positively related to probability of DBP violations (8.1%) (table S14). This is consistent with DBP rule violations occurring in low-income, suburban systems [16], relative to the geography of inorganics which predominate in rural systems sourced by groundwater. Surface Water Treatment rule violations were most clearly explained by socioeconomic factors (figure 6(e)). There was little consistency in the ranking of social parameters for the remaining drinking water quality violation rules with all three themes (socioeconomic, race + language, demographic) contributing to violations (figure 6, table S17).

3.4. Recurrence of violations and relationship to modified social vulnerability index

Recurrence of any HB violation was also strongly linked to mSVI, particularly for certain rules (figure 7), indicating that CWSs with the most persistent, systemic violations are also the most socially vulnerable. Median mSVI increased with recurrence of any HB violation based on quarterly data for 2018–2020 ($R = 0.73$; figure 7(a)). All correlation coefficients in this study are highly significant, with p values <0.0001. Moderate to strong relationships between median mSVI of violating systems and number of quarters in violation were found for inorganic contaminants (arsenic, $R = 0.73$; nitrate, $R = 0.44$) and treatment issues (DBP rules, $R = 0.54$) (figures 7(b)–(d)) despite the vastly different formation and mobilization mechanisms of these contaminants. Analysis of initial and recurring violations of selected rules reveals that systems with lower socioeconomic status and higher proportions of minority groups are more likely to experience initial and recurring drinking water violations [1].

In contrast, violations of most of the other rules, including Surface Water Treatment, Ground Water, Lead & Copper, and Revised Total Coliform rules did not extend beyond the first few quarters, suggesting that violations of treatment technique rules may be

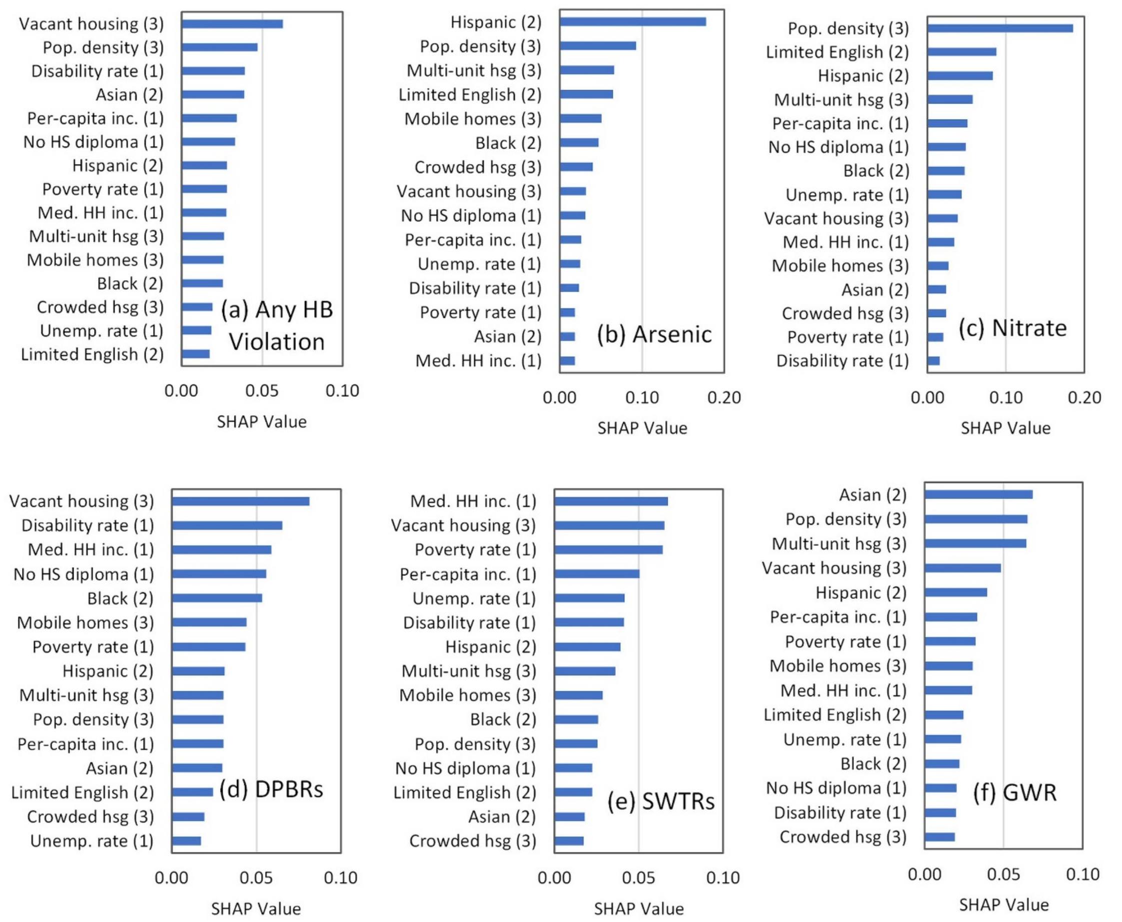

Figure 6. SHAP values showing relative importance of modified social vulnerability index (mSVI) parameters to (a) any HB violation, (b) DBPRs, (c) arsenic, (d) nitrate, (e) surface water treatment rules, and (f) GWR violations, based on 2018–2020 SDWA violations. SHAP values essentially represent the change in percent likelihood of incurring a violation. Numbers in parentheses refer to three SVI themes: 1: Socioeconomic Status; 2: Minority status & language, 3: Housing & population density; (figure S1). Data are provided in table S9(e). A more complete set of plots is provided in figure S11. Pop. density: population density; Per-capita inc.: Per-capita income; No HS diploma: No High-School diploma; Med HH Inc.: Median Household Income; Multi-unit hsg: Multi-unit housing; Crowded hsg: Crowded housing; Unemp. rate: Unemployment rate.

more readily addressed across all CWSs, regardless of social vulnerability status (table S21).

3.5. Implications for defining disadvantaged communities

The new federal drinking water infrastructure funding requires states to allocate ⩾49% of the funding to DACs [23]. States have substantial discretion in defining DACs, resulting in large variations in DAC definitions across the US. Currently, 49 of 51 states incorporate MHI in their definitions with eight states relying solely on MHI or a similar metric. New York and Maryland are the only two states that currently incorporate EJ criteria, including unemployment, population decline, and race [23]. Results of this analysis suggest that a broader definition of DACs, beyond MHI, should be considered as mSVI captures 3× more of the affected population than MHI considering the most vulnerable 25% of mSVI compared to the lowest 25% of MHI (figure 8).

Our analysis highlights the broader dimensions of social vulnerability, beyond pure economic metrics, which explain linkages with noncompliant CWSs. To date, both state and federal environmental justice programs employ a single metric, primarily based on economic factors, in defining social vulnerability [22, 32]. Additional social parameters with existing data sources at the local level should be considered, including rurality, race and language, population trends, and housing type and tenure when defining DACs, to better capture disparities in access to safe drinking water quality. In addition, the emphasis on drinking water quality in this study suggests that environmental justice tools should expand their consideration of drinking water quality issues beyond point sources and include naturally occurring geogenic and non-point source anthropogenic contaminants [16, 17], as applied in the California Community Water Center tool [42]. While there are inherent limitations to using social vulnerability indices, policy makers and states currently use less rigorous metrics to define DACs and distribute funds. Accordingly, we propose the potential use of mSVI for drinking water quality violations, which is informed by literature and

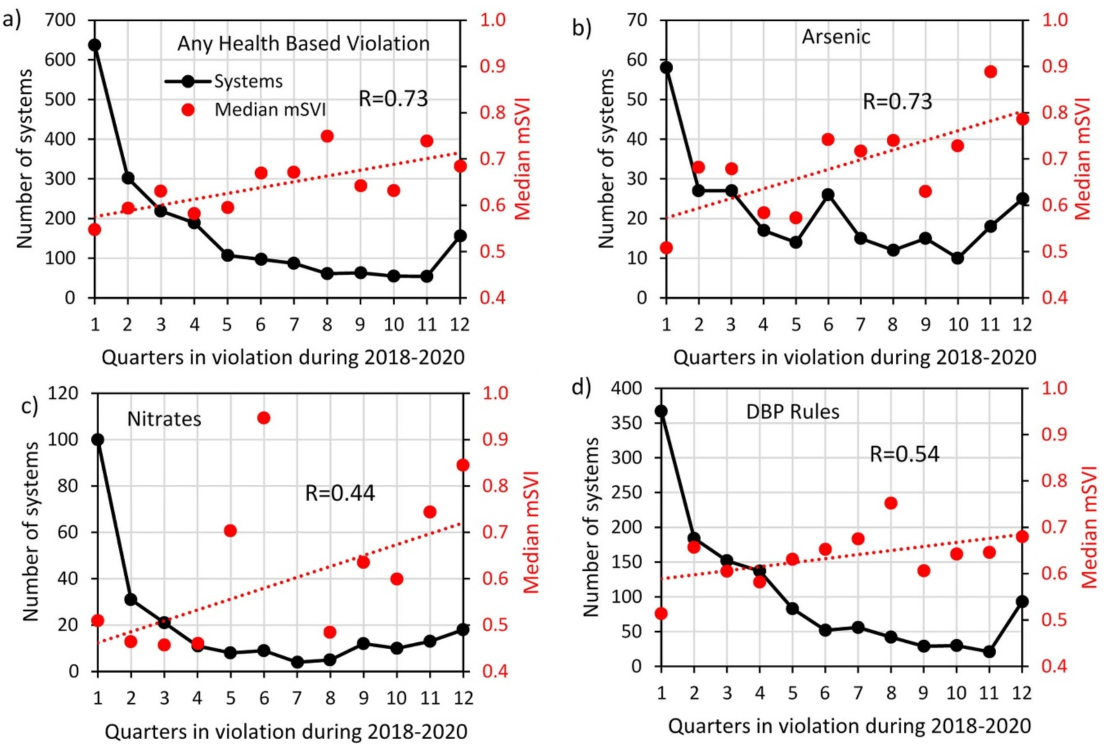

Figure 7. Percent of CWSs with (a) any HB violations, (b) Arsenic violations, (c) Nitrates violations, and (d) DPBR violations versus the number of quarters those systems had violations during 2018–2020 (5041 systems total). The median mSVI of CWSs in each group is also shown. All plots and data are provided in table S18e and figure S24. mSVI is normalized for CONUS using a percent rank, resulting in a median value of 0.5 for all CWSs. All R values are statistically significant, with p values <0.0001.

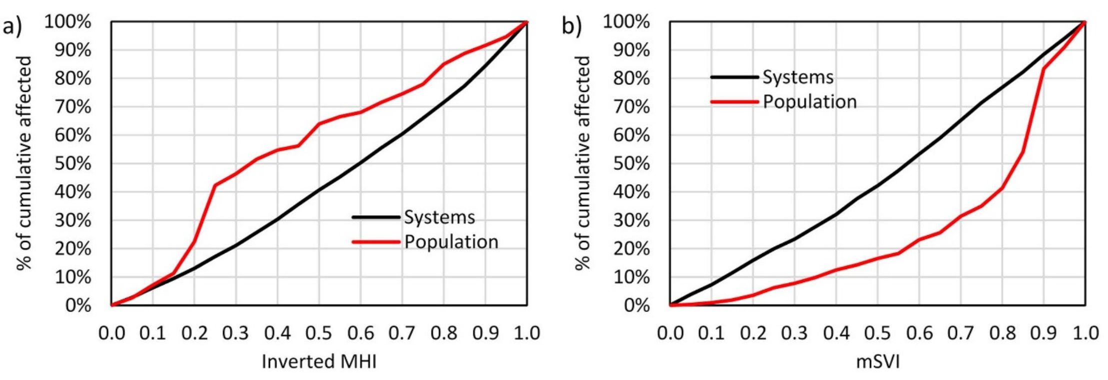

Figure 8. Comparison between (a) inverted median household income (MHI) scores and (b) modified social vulnerability index (mSVI) in assessingvulnerability of CWSs with any health-based violation. Inverted indicates the complimentary inversion (i.e. $1 - X$) of the MHI score such that the lowest MHI dollar values are represented by the highest scores and vice-versa. See discussion in supplementary material section 2.9. Data are provided in table S4(b).

data analytics, to better capture inequities in SDWA violation incidence and recurrence in the US. The proposed mSVI is currently exploratory and would need to be validated/tested more rigorously to lean into a recommendation on wider usage beyond this study.

Continual expansion of the database of CWS service area boundaries [27] will provide more comprehensive coverage to match to mSVI across the CONUS to better understand state-specific demographics and socioeconomic drivers of SDWA violations. Tribal systems should also be considered in much more detail in efforts to conduct more refined spatial analyses. While we evaluated HB violations of tribal CWSs, many tribal regions lack access to any water infrastructure, highlighting the multiple dimensions of water inequities [3], beyond drinking water quality.

3.6. Implications for achieving drinking water compliance

This study has important implications for solving drinking water quality issues across the U.S., especially in DACs. Incidents and recurrence of drinking water quality violations are strongly linked to social vulnerability.

One solution to address water quality violations in socially vulnerable communities is to prevent source water contamination [28]; however, this may have limited relevance to tackling pervasive naturally occurring geogenic contaminants (e.g. arsenic, radionuclides) that impact socially vulnerable communities in the Southwest and Southcentral US [16]. Similarly, nitrate contamination is generally a nonpoint source contaminant that remains in groundwater for decades and is linked to agriculture in much of the Southwest and Southcentral US [35]. It is difficult for CWSs to address pervasive arsenic [43] and nitrate [35] contamination in groundwater given the lack of surface water availability to blend with groundwater or provide an alternative water source in the overexploited, semiarid Southwest and Southcentral US. The only option is often installation of treatment systems; however, this is extremely challenging for small rural systems serving marginalized communities with limited financial, managerial, and technical capacity. Point-of-use and point-of-entry systems are and must increasingly be considered as an alternative to centralized treatment systems [44].

In addition, potential merger, regionalization or CWS consolidation partners may bias away from systems serving low-income and minority communities due to high transaction costs [45], highlighting the need to ensure that DAC definitions effectively prioritize the most vulnerable communities for funding and assistance efforts. We show that recurrence of violations is a critical issue; however, the new federal infrastructure funding does not allow funding for ongoing operation and maintenance of CWSs, which is paramount to ensuring long-term system compliance. More promisingly, the federal EPA's new Water System Restructuring Assessment Rule will require primacy agencies (states and tribes) to assess the feasibility of restructuring for noncompliant systems and provides incentives for implementation [46].

Ultimately, many of the challenges DACs face in achieving and maintaining compliance result from limited financial, managerial, and technical capacity. Many lack the staff to apply for funding, the expertise to operate infrastructure, and the governance structures to ensure inclusive representation of community needs. Providing more technical assistance and management options to these systems and to community-based organizations working with DACs is essential to enhance these programs and overcome recurrent drinking water violations which continue to impair public health in the US.

4. Conclusions

The mSVI developed in this study provides a novel index targeted specifically to drinking water quality violations, matched to actual CWS locations. We show that the incidence of drinking water violations disproportionately impacts socially vulnerable communities, with ∼70% of the population served ranking in the high mSVI category, and many different social parameters, beyond simple income or socioeconomic metrics, linked to different drinking water quality violations. However, current federal environmental justice tools and existing DAC definitions often place emphasis on socioeconomic factors, potentially neglecting key, data-available aspects of social vulnerability related to equitable access to safe drinking water.

Additionally, mSVI is strongly linked to recurrence of select drinking water quality violations (i.e. the ability of systems to recover from violations over time), including any HB, arsenic, and DBP violations. The recurrence of inorganic violations reflects the pervasive distribution of naturally occurring arsenic and anthropogenic nitrate in the southwest and southcentral US, requiring treatment systems and the limited ability of high mSVI communities to install and maintain such treatment systems. This study has important implications for future studies to address drinking water quality issues via certain treatment options for specific rules and emphasizing the importance of additional operational funding to maintain treatment systems where there are recurrent violations. Our detailed analysis of the linkages of drinking water quality violations to social vulnerability can help inform guidance for effectively distributing infrastructure funding and designing interventions to ensure more equitable drinking water quality nationally.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

We acknowledge review comments from Dr James Hogan (EPA Headquarters) and Dr Tom Mueller (Univ. of Oklahoma) that substantially improved the manuscript.

Conflict of interest

The authors declare no conflict of interest.

Ethical statement

The research was conducted in accordance with the principles embodied in the Declaration of Helsinki and in accordance with local statutory requirements.

Funding source

This study was funded by the US EPA through the Texas Commission on Environmental Quality, the

Fisher Endowment, the Jackson Endowment, and the State of Texas Advanced Resources Recovery.

ORCID iDs

Bridget R Scanlon https://orcid.org/0000-0002-1234-4199
Sarah Fakhreddine https://orcid.org/0000-0003-3583-9327
Qian Yang https://orcid.org/0000-0002-2266-7231

References

[1] McDonald Y J and Jones N E 2018 Drinking water violations and environmental justice in the United States, 2011–2015 *Am. J. Public Health* **108** 1401–7
[2] Nigra A E, Chen Q, Chillrud S N, Wang L, Harvey D, Mailloux B, Factor-Litvak P and Navas-Acien A 2020 Inequalities in public water arsenic concentrations in counties and community water systems across the United States, 2006–2011 *Environ. Health Perspect.* **128** 127001
[3] Mueller J T and Gasteyer S 2021 The widespread and unjust drinking water and clean water crisis in the United States *Nat. Commun.* **12** 3544
[4] Johnson H Jr 2022 Challenges of an aging water system: the Jackson water crisis—a research commentary *Online Journal of Rural and Urban Research: Mississippi Urban Research Center: Special Issue: Implications of the 2021 Jackson Water Crisis: Past, Present, and Future* pp 5–21
[5] Ko J Y 2022 An explanation of the 2021 Jackson water crisis and policy suggestions for sustainable water infrastructure in Jackson, Mississippi—a research commentary *Online Journal of Rural and Urban Research: Mississippi Urban Research Center: Special Issue: Implications of the 2021 Jackson Water Crisis: Past, Present, and Future* pp 5–71
[6] Simonson A and Wolfe E 2022 EPA determines water in Jackson, Mississippi, is safe to drink two months after treatment plant failure (available at: www.cnn.com/2022/10/31/us/jackson-mississippi-water-safe-to-drink-epa)
[7] Morckel V and Terzano K 2019 Legacy city residents' lack of trust in their governments: an examination of Flint, Michigan residents' trust at the height of the water crisis *J. Urban Aff.* **41** 585–601
[8] Rosinger A Y, Patel A I and Weaks F 2021 Examining recent trends in the racial disparity gap in tap water consumption: NHANES 2011–2018 *Public Health Nutr.* **25** 1–7
[9] Javidi A and Pierce G 2018 US households' perception of drinking water as unsafe and its consequences: examining alternative choices to the tap *Water Resour. Res.* **54** 6100–13
[10] ASCE Report card for America's infrastructure, drinking water *American Society of Civil Engineering* (available at: www.infrastructurereportcard2021org/cat-item/drinking_water/) (Accessed 22 August 2021)
[11] AWWA 2012 Buried no longer: confronting America's water infrastructure challenge *Am. Water Works Assoc.* p 37
[12] Board C W Human right to water portal 2021 (available at: www.waterboards.ca.gov/water_issues/programs/hr2w/#:∼:text=Human%20Right%20to%20Water%20%7C%20California%20State%20Water%20Resources%20Control%20Board&text=On%20September%2025%2C%202012%2C%20Governor,the%20human%20right%20to%20water) (Accessed 17 July 2021)
[13] Rivera-Juaristi F J and Schwartz B 2015 The human right to water in the United States *Santa Clara Law, Intl. Human RIghts Clinic* vol 29 (available at: https://law.scu.edu/wp-content/uploads/150915_IACHR-Water-Rts-Questionnaire_United-States_Santa-Clara.pdf)
[14] Congress H R 2021 *3684–117th Congress (2021–2022)—Infrastructure Investment and Jobs Act* (available at: www.congress.gov/bill/117th-congress/house-bill/3684/text)
[15] Fox R 2022 Implementation of the clean water and drinking water state revolving fund provisions of the Bipartisan Infrastructure Law *EPA Memorandum* vol 56 (available at: www.epa.gov/system/files/documents/2022-03/combined_srf-implementation-memo_final_03.2022.pdf) (Accessed 8 March 2022)
[16] Scanlon B R, Fakhreddine S, Reedy R C, Yang Q and Malito J G 2022 Drivers of spatiotemporal variability in drinking water quality in the United States *Environ. Sci. Technol.* **56** 12965–74
[17] Belitz K, Fram M S, Lindsey B D, Stackelberg P E, Bexfield L M, Johnson T D, Jurgens B C, Kingsbury J A, McMahon P B and Dubrovsky N M 2022 Quality of groundwater used for public supply in the continental United States: a comprehensive assessment *ACS ES&T Water* **2** 2645–56
[18] EPA 2019 Stage 2 disinfectants and disinfection byproducts rule (DBPR) and consecutive in-depth analysis *EPA Report* vol 35
[19] CDC 2021 Social Vulnerability Index (CDC SVI): a tool to identify socially vulnerable communities *Centers for Disease Control and Prevention, Agency for Toxic Substances and Disease Registry, CDC-ASTDR Fact Sheet* (available at: www.atsdr.cdc.gov/placeandhealth/svi/fact_sheet/pdf/SVI_FactSheet_v10152020-H.pdf)
[20] Flanagan B E, Hallisey E J, Adams E and Lavery A 2018 Measuring community vulnerability to natural and anthropogenic hazards: the centers for disease control and prevention's Social Vulnerability Index *J. Environ. Health* **80** 34–36
[21] Council on Environmental Quality 2022 Climate and economic justice screening tool *Technical Support Document, Ver* vol 1 p 36
[22] EPA 2019 *EJSCREEN, Environmental Justice Mapping and Screening Tool Technical Documentation* (U. S. Environmental Protection Agency, Office of Policy) p 115
[23] EPA 2022 Drinking water state revolving fund, disadvantaged community definitions: a reference for states *U.S. EPA Report, EPA 810-R-22-002* vol 41
[24] Allaire M, Wu H W and Lall U 2018 National trends in drinking water quality violations *Proc. Natl Acad. Sci. USA* **115** 2078–83
[25] Switzer D and Teodoro M P 2017 The color of drinking water: class, race, ethnicity, and Safe Drinking Water Act compliance *J. Am. Water Works Assoc.* **109** 40–45
[26] Pierce G, Lai L and DeShazo J R 2019 Identifying and addressing drinking water system sprawl, its consequences, and the opportunity for planners? Intervention: evidence from Los Angeles County *J. Environ. Plan.* **62** 2080–100
[27] McDonald Y J, Anderson K M, Caballero M D, Ding K J, Fisher D H, Morkel C P and Hill E L 2022 A systematic review of geospatial representation of United States community water systems *AWWA Water Sci.* **4** e1266
[28] Fedinick K P, Taylor S and Roberts M 2019 Watered down justice *Natural Resources Defense Council, Coming Clean, and Environmental Justice Health Alliance Rept.* vol 51
[29] Michielssen S, Vedrin M C and Guikema S D 2020 Trends in microbiological drinking water quality violations across the United States *Environ. Sci. Water Res. Technol.* **6** 3091–105
[30] Pennino M J, Compton J E and Leibowitz S G 2017 Trends in drinking water nitrate violations across the United States *Environ. Sci. Technol.* **51** 13450–60
[31] Pennino M J, Leibowitz S G, Compton J E, Hill R A and Sabo R D 2020 Patterns and predictions of drinking water nitrate violations across the conterminous United States *Sci. Total Environ.* **722** 137661
[32] WHEJAC 2021 White house environmental justice advisory council: final recommendations: justice40, climate and economic justice screening tool & executive order 12898 revisions *WHEJAC Final Report Executive Order* **14008** 91

[33] Hughes S 2022 A multidimensional approach to evaluating the vulnerability of drinking water systems *J. Environ. Policy Plan.* **24** 210–26

[34] Stillo F and MacDonald Gibson J 2017 Exposure to contaminated drinking water and health disparities in North Carolina *Am. J. Public Health* **107** 180–5

[35] Harter T *et al* 2012 Addressing nitrate in California's drinking water with a focus on Tulare Lake Basin and Salinas valley groundwater *Report for the State Water Resources Control Board Report to the Legislature, Center for Watershed Sciences, University of California Davis* vol 78 (available at: http://groundwaternitrate.ucdavis.edu)

[36] SimpleLab EPIC 2022 U.S. Community water systems service boundaries, V2.0.0, HydroShare (available at: www.hydroshare.org/resource/20b908d73a784fc1a097a3b3f2b58bfb)

[37] ECHO 2022 Safe Drinking Water Act data download summary and data element dictionary *EPA ECHO Enforcement and Compliance History Online, SDWA Data Download Summary and Data Element Dictionary* (Accessed 15 July 2022)

[38] Liu X, Wu J and Zhou Z 2009 Exploratory undersampling for class-imbalance learning *IEEE Trans. Syst. Man Cybern.* B **39** 539–50

[39] Lundberg S M *et al* 2018 Explainable machine-learning predictions for the prevention of hypoxaemia during surgery *Nat. Biomed. Eng.* **2** 749–60

[40] Lundberg S M, Erion G, Chen H, DeGrave A, Prutkin J M, Nair B, Katz R, Himmelfarb J, Bansal N and Lee S-I 2020 From local explanations to global understanding with explainable AI for trees *Nat. Mach. Intell.* **2** 56–67

[41] Teodoro M P, Haider M and Switzer D U S 2016 Environmental policy implementation on tribal lands: trust, neglect, and justice *Policy Stud. J.* **46** 37–59

[42] Renteria A, Lupercio A V, Okita D, Ores D, Lukacs H, Firestone L, Grazier N G, Jensen R and Anda S D Community water center drinking water tool (available at: https://drinkingwatertool.communitywatercenter.org/ca-water/) (Accessed 31 January 2023)

[43] Ayotte J D, Medalie L, Qi S L, Backer L C and Nolan B T 2017 Estimating the high-arsenic domestic-well population in the conterminous United States *Environ. Sci. Technol.* **51** 12443–54

[44] Wu J, Cao M, Tong D, Finkelstein Z and Hoek E M V 2021 A critical review of point-of-use drinking water treatment in the United States *npj Clean Water* **4** 40

[45] Lee M Y A and Braden J B 2008 Examining mergers in small CWSs: the role of regulatory compliance *J. AWWA* **100** 58–67

[46] EPA 2019 America's water infrastructure act: promoting water system compliance and long-term capacity via the water system restructuring rule *EPA #810-F-19-009, 2* (available at: www.epa.gov/sites/default/files/2019-07/documents/awia-_restructuring_rule_factsheet.pdf)